
Cokriging and multivariate kriging methods based on data of a functional random field¹

Métodos de Cokriging y kriging multivariado basados en datos de un campo aleatorio funcional

Ramón Giraldo^a
rgiraldoh@unal.edu.co

Pedro Delicado^b
pedro.delicado@upc.edu

Jorge Mateu^c
mateu@mat.uji.es

Abstract

Kriging and cokriging and their several related versions are techniques widely known and used in spatial data analysis. However, when the spatial data are functions a bridge between functional data analysis and geostatistics has to be built. We give an overview to cokriging analysis and multivariable spatial prediction to the case where the observations at each sampling location consist of samples of random functions. We extend multivariable geostatistical methods to the functional context. Our cokriging method predicts one variable at a time as in a classical multivariable sense, but considering as auxiliary information curves instead of vectors. We also give an extension of multivariable kriging to the functional context where is defined a predictor of a whole curve based on samples of curves located at a neighborhood of the prediction site. In both cases a non-parametric approach based on basis function expansion is used to estimate the parameters, and we prove that both proposals coincide when using such an approach. A linear model of coregionalization is used to define the spatial dependence among the coefficients of the basis functions, and therefore for estimating the functional parameters. As an illustration the methodological proposals are applied to analyze two real data sets corresponding to average daily temperatures measured at 35 weather stations located in the Canadian Maritime Provinces, and penetration resistance data collected at 32 sampling sites of an experimental plot.

Keywords: basis functions; Cross-validation, functional linear model; linear model of coregionalization, multivariable cokriging.

¹DOI: <http://dx.doi.org/10.15332/s2027-3355.2017.0002.07>

Giraldo, R., Delicado, P. and Mateu, J. (2017) Cokriging and multivariate kriging methods based on data of a functional random field. *Comunicaciones en Estadística*, **10**(2), 315-344.

^aStatistics Department, Universidad Nacional de Colombia

^bDepartament d'Estadística i Investigació Operativa, Universitat Politècnica de Catalunya

^cDepartment of Mathematics, Universitat Jaume I

Resumen

Kriging y cokriging y sus versiones relacionadas son técnicas ampliamente conocidas y usadas en análisis de datos espaciales. Sin embargo cuando los datos son curvas se requiere un puente entre análisis de datos funcionales y geoestadística. Aquí se da una revisión sobre el uso de cokriging y de kriging multivariado para el caso en que las observaciones en cada sitio de muestreo corresponden a muestras de funciones aleatorias. Se extiende la geoestadística multivariada al contexto funcional. El método cokriging propuesto permite predecir una variable en un periodo de tiempo como en el sentido multivariado, pero considerando como variables auxiliares curvas en vez de vectores. También se muestra como extender el kriging multivariable al caso funcional definiendo un predictor de una curva completa basado en la información de curvas en sitios cercanos. En las dos propuestas se usan métodos no paramétricos de suavizado por base de funciones. Se comprueba que las dos aproximaciones consideradas coinciden. Se emplea un modelo lineal de correogionalización (MLC) para definir la dependencia entre los coeficientes resultantes del ajuste de las bases de funciones y por consiguiente para la estimación de los parámetros funcionales. A manera de ilustración la metodología propuesta es aplicada al análisis de dos conjuntos de datos reales correspondientes por una lado a temperaturas diarias medidas en 35 estaciones meteorológicas de las provincias marítimas de Canada y por el otro a datos de resistencia mecánica a la penetración colectados en 32 puntos de muestreo en una parcela experimental.

Palabras clave: base de funciones, cokriging multivariado, modelo lineal de correogionalización, modelo lineal funcional, validación cruzada.

1 Introduction

In recent years there has been an increasing interest in modeling functional data. In environmental studies this type of data often arise when measurements are recorded at a fine grid of temporal instants over a period of time. Statistical methods for analyzing this type of data define a new branch of statistics called Functional Data Analysis (FDA). Exploratory data analysis (Ramsay & Silverman 2005), regression (Cardot et al. 2007), analysis of variance (Cuevas et al. 2004), non-parametric analysis (Ferraty & Vieu 2006) and multivariate techniques (Ferraty & Vieu 2003) are standard statistical procedures adapted to the functional setting. An overview of statistical methods for analyzing functional data is shown in Ramsay & Silverman (2005) and recent developments in this field are given in special issues of several journals (González-Manteiga & Vieu 2007).

In this paper we focus on spatially correlated functional data, and particularly in modeling curves collected in sites of a region with spatial continuity. In spatial statistics, and in particular in geostatistics, both cokriging analysis (Bogaert 1996) and multivariable kriging (Ver Hoef & Barry 1998) are used for modeling observations of vector-valued random fields. Here we adapt these methodologies to the functional context. We propose a cokriging predictor (*cokriging based on curves*)

for performing prediction at a particular point of the curve (as in the cokriging multivariable sense), but considering as auxiliary information samples of curves instead of observations of random vectors. Likewise, we extend the multivariable kriging from random vectors to the functional context by defining the *functional kriging total model* predictor which allows to predict a whole curve at an unvisited site by using as information the curves sampled in the neighborhood of the prediction site. In both cases we give a non-parametric solution based on basis functions. we also prove that both approaches are equivalent for each particular point of the curve. A distinctive feature between these two methodologies is given in terms of their prediction variances. The estimated prediction variance of the functional kriging total model can be used as a global measure of uncertainty in the prediction of a whole curve, whereas the prediction variance of the cokriging based on curves method can be used in a classical sense, that is, we can calculate point-wise confidence intervals for the prediction.

The problem of functional kriging for predicting a whole curve is also studied by Nerini et al. (2010). Their proposal (called cokriging for functional data) is based on the use of orthonormal basis functions. This is justified by the fact that it allows to find a simple expression of the minimization problem and the functional kriging problem reduces to a standard multivariate kriging on the coefficients. That approach limits the use of basis such as B-spline basis functions which are frequently used in the field of functional data analysis (Ramsay & Silverman 2005). In this sense, our functional kriging predictor is more general because orthogonality is not a required condition to give a solution to the minimization problem. In addition, Nerini et al. (2010), do not consider a predictor to perform predictions at particular points of the curve, as our proposal for cokriging based on curves does. With their functional kriging and cokriging functional predictors it is only possible to estimate an integrated prediction variance. One important advantage in our functional proposals is that, in addition to the prediction for a particular point of the curve, we can also estimate the prediction variance at this point of the curve.

The problem of spatial prediction of curves in a geostatistical context has been considered from several points of view. Goulard & Voltz (1993) is a pioneer work in this topic. They propose three geostatistical approaches to predict curves: a curve kriging approach and two multivariable approaches based on cokriging on either discrete data or coefficients of the parametric models that have been fitted to the observed curves. Giraldo et al. (2011), propose a non-parametric approach for solving the first approach considered by Goulard & Voltz (1993). The predictor in the first proposal of Goulard & Voltz (1993) as well as that considered by Giraldo et al. (2011), has the same form as the classical ordinary kriging predictor (Cressie 1993), but considering curves instead of one-dimensional data, that is, each curve is weighted by an scalar parameter. Giraldo et al. (2010), solve the problem of spatial prediction of functional data by weighting each observed curve by a functional parameter. This approach is a hybrid between ordinary kriging and the functional linear concurrent (*point-wise*) model such as shown in Ramsay & Silverman (2005). The methodologies proposed in this paper follow the line

of Giraldo et al. (2010), in the sense that each observed curve is weighted by a functional parameter. However, here the flexibility increases because double indexed functional parameters are considered and estimated. Now, each curve is weighted by a functional parameter for the prediction at each time. This modeling approach follows the basic philosophy of the functional linear model for functional response, for which a bivariate regression coefficient function must be estimated (Malfait & Ramsay 2003).

The methodology proposed is illustrated by means of applications to two real datasets. A good knowledge of the spatial and temporal patterns of meteorological and climate variables is often required when dealing with environmental problems. These variables are often measured in a fine temporal grid and over a set of spatial locations. What we observe can be thus considered a sample of a continuous function, and the spatial functional methodology is the right scientific context. Here we show how the predictors proposed can be used in modeling a data set consisting of daily temperature measurements recorded at 35 weather stations of the Canadian Maritime Provinces. In addition we apply the methodology to an agronomic data set corresponding to penetration resistance data recorded at 35 sites of an experimental plot at the National University of Colombia.

The remainder of the paper is organized as follows. Section 2 presents an overview of cokriging and multivariable kriging. Section 3 introduces the cokriging predictor based on curves and the corresponding parameter estimation. In Section 4 an extension of multivariable kriging to the functional context is shown and its equivalence with cokriging based on curves is established. Applications of the proposed methodology to the Canadian Maritime Provinces temperature data, and the penetration resistance data set are given in Section 5. In the first example Fourier basis functions and B-splines basis functions are used. In the second only a B-splines basis functions is used taking into account that the data are not periodic. The paper ends with a brief discussion and suggestions for further research.

2 Cokriging and multivariable spatial prediction

In this section we show the basics of cokriging (Myers 1982) and multivariable spatial prediction (Ver Hoef & Cressie 1993). As in Ver Hoef & Barry (1998) we use the term cokriging to mean prediction of a single random variable, and the term multivariable spatial prediction when predicting a vector of random variables. Let $\{\mathbf{Z}(s) = (Z_1(s), \dots, Z_m(s)) : s \in D\}$ be a multivariable spatial process defined over a domain $D \subset \mathbb{R}^d$. In practice usually $d = 2$, with $s = (x, y)$ corresponding to the geographical coordinates. We now consider the model

$$\mathbf{Z}(s) = \boldsymbol{\mu}(s) + \boldsymbol{\epsilon}(s),$$

where $\boldsymbol{\mu}(s)$ is a mean vector and $\boldsymbol{\epsilon}(s)$ a random vector with expected value $\mathbb{E}(\boldsymbol{\epsilon}(s)) = \mathbf{0}$. It is assumed that the process is stationary, that is, the mean vector is considered constant for all $s \in D$, and the variance (covariance), cross-covariance and cross-variogram functions depend only on the separation vector

\mathbf{h} , and not on locations. Let $\{\mathbf{Z}(s_i) = (Z_1(s_i), \dots, Z_m(s_i)), i = 1, \dots, n\}$ be a realization of a multivariable spatial process and $Z_k(s_0), k = 1, \dots, m$ a random variable at $s_0 \in D$. We use the following notation:

$2\gamma_{lq}(s_i, s_j) = \mathbb{V}(Z_l(s_i) - Z_q(s_j)), l, q = 1, \dots, m, i, j = 1, \dots, n$, where \mathbb{V} stands for the variance.

$$\boldsymbol{\gamma}_{lk}^T = (\gamma_{lk}(s_1, s_0), \dots, \gamma_{lk}(s_n, s_0))$$

$$\boldsymbol{\Gamma}_{lq} = \begin{pmatrix} \gamma_{lq}(s_1, s_1) & \cdots & \gamma_{lq}(s_1, s_n) \\ \vdots & \ddots & \vdots \\ \gamma_{lq}(s_n, s_1) & \cdots & \gamma_{lq}(s_n, s_n) \end{pmatrix}$$

The cokriging predictor of the random variable $Z_k(s_0)$ based on the realization $\mathbf{Z}(s_i)$ is given by

$$\begin{aligned} \hat{Z}_k(s_0) &= \sum_{j=1}^m \lambda_{1j}^k Z_j(s_1) + \cdots + \sum_{j=1}^m \lambda_{nj}^k Z_j(s_n) \\ &= \sum_{i=1}^n \sum_{j=1}^m \lambda_{ij}^k Z_j(s_i). \end{aligned} \quad (1)$$

The predictor (1) is unbiased if $\sum_{i=1}^n \lambda_{ik}^k = 1$ and $\sum_{i=1}^n \lambda_{ij}^k = 0$ for all $j \neq k, j = 1, \dots, m$. Using the method of Lagrange multipliers to minimize the mean squared prediction error, $\mathbb{E}(\hat{Z}_k(s_0) - Z_k(s_0))^2$, subject to the unbiasedness constraints gives the cokriging system of equations, which in matrix notation can be expressed by

$$\mathbf{C}\boldsymbol{\lambda}^k = \mathbf{c}^k,$$

with

$$\mathbf{C} = \begin{pmatrix} \boldsymbol{\Gamma}_{11} & \cdots & \boldsymbol{\Gamma}_{1k} & \cdots & \boldsymbol{\Gamma}_{1m} & \mathbf{1} & \cdots & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \boldsymbol{\Gamma}_{k1} & \cdots & \boldsymbol{\Gamma}_{kk} & \cdots & \boldsymbol{\Gamma}_{km} & \mathbf{0} & \cdots & \mathbf{1} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \boldsymbol{\Gamma}_{m1} & \cdots & \boldsymbol{\Gamma}_{m2} & \cdots & \boldsymbol{\Gamma}_{mm} & \mathbf{0} & \cdots & \mathbf{0} & \cdots & \mathbf{1} \\ \mathbf{1}^T & \cdots & \mathbf{0}^T & \cdots & \mathbf{0}^T & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \mathbf{0}^T & \cdots & \mathbf{1}^T & \cdots & \mathbf{0}^T & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \mathbf{0}^T & \cdots & \mathbf{0}^T & \cdots & \mathbf{1}^T & 0 & \cdots & 0 & \cdots & 0 \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Gamma} & \mathbf{X} \\ \mathbf{X}^T & \mathbf{0}^* \end{pmatrix}, \quad (2)$$

$$\boldsymbol{\lambda}^k = \begin{pmatrix} \lambda_1^k \\ \vdots \\ \lambda_k^k \\ \vdots \\ \lambda_m^k \\ \delta_1 \\ \vdots \\ \delta_k \\ \vdots \\ \delta_m \end{pmatrix}, \mathbf{c}^k = \begin{pmatrix} \gamma_1^k \\ \vdots \\ \gamma_k^k \\ \vdots \\ \gamma_m^k \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix},$$

where $(\boldsymbol{\Gamma}_{ij})_{(n \times n)}$, $\mathbf{1} = (1, \dots, 1)_{(n \times 1)}^T$, $\mathbf{0} = (0, \dots, 0)_{(n \times 1)}^T$, $(\boldsymbol{\Gamma})_{(m \times n) \times (n \times m)}$, $(\mathbf{X})_{(n \times m) \times m}$, $(\mathbf{0}^*)_{(m \times m)}$, $\boldsymbol{\lambda}_j^k = (\lambda_{1j}^k, \dots, \lambda_{nj}^k)^T$, and $\boldsymbol{\gamma}_j^k = (\gamma_{1j}^k, \dots, \gamma_{nj}^k)^T$, for all $i, j = 1, \dots, m$.

In multivariable spatial prediction (Ver Hoef & Cressie 1993, Ver Hoef & Barry 1998) all the m variables are predicted simultaneously at s_0 . In this case the predictor kriging is given by

$$\begin{pmatrix} \hat{Z}_1(s_0) \\ \vdots \\ \hat{Z}_m(s_0) \end{pmatrix} = \begin{pmatrix} \lambda_{11}^1 & \cdots & \lambda_{1m}^1 & \cdots & \lambda_{n1}^1 & \cdots & \lambda_{nm}^1 \\ \vdots & \ddots & \vdots & & \vdots & \ddots & \vdots \\ \lambda_{11}^m & \cdots & \lambda_{1m}^m & \cdots & \lambda_{n1}^m & \cdots & \lambda_{nm}^m \end{pmatrix} \begin{pmatrix} Z_1(s_1) \\ \vdots \\ Z_m(s_1) \\ \vdots \\ Z_1(s_n) \\ \vdots \\ Z_m(s_n) \end{pmatrix}, \quad (3)$$

and the parameters are estimated by minimizing the matrix $\mathbb{V}(\hat{\mathbf{Z}}(s_0) - \mathbf{Z}(s_0))$. The solution is then obtained by solving the system (Ver Hoef & Barry 1998)

$$\begin{pmatrix} \boldsymbol{\Gamma} & \mathbf{X} \\ \mathbf{X}^T & \mathbf{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{\Lambda} \\ \boldsymbol{\Delta} \end{pmatrix} = \begin{pmatrix} \mathbf{G} \\ \mathbf{I} \end{pmatrix},$$

where $\boldsymbol{\Gamma}$ and \mathbf{X} are defined as in (2), $\boldsymbol{\Lambda}$ is the matrix of parameters, $\boldsymbol{\Delta}$ is a diagonal matrix of Lagrange multipliers, \mathbf{I} is an identity matrix and

$$\mathbf{G} = \begin{pmatrix} \gamma_1^1 & \gamma_1^2 & \cdots & \gamma_1^m \\ \gamma_2^1 & \gamma_2^2 & \cdots & \gamma_2^m \\ \vdots & & \ddots & \vdots \\ \gamma_m^1 & \gamma_m^2 & \cdots & \gamma_m^m \end{pmatrix}.$$

Cokriging could be used for predicting simultaneously all m variables by cokriging each variable, one at a time. The cokriging prediction for one variable at a time coincides with the prediction of that variable obtained by multivariable spatial

prediction (Ver Hoef & Cressie 1993). However, this coincidence depends on the criteria used in multivariable prediction. If for example criteria takes into account cross-correlation between variables using the Mahalanobis metric the results need not be similar (see details in Ver Hoef & Cressie (1993)). Proposition 1 in this paper clarifies this fact. The difference between both approaches is given by their prediction variances. With cokriging we obtain a prediction variance for each univariate variable. In multivariable spatial prediction, in addition to the univariate prediction variances, is possible to estimate a multidimensional prediction region with its long axis oriented toward regions where the predicted variables tend to co-vary (Ver Hoef & Cressie 1993).

3 Cokriging based on curves

Let $\{\chi_s(t), t \in T, s \in D \subset \mathbb{R}^d\}$ be a random function defined on some compact set T of \mathbb{R} . Suppose we observe a sample of curves $\chi_{s_1}(t), \dots, \chi_{s_n}(t)$ defined for $t \in T$, $s_i \in D, i = 1, \dots, n$. We assume that these curves belong to a separable Hilbert space \mathbf{H} of square integrable functions defined on T . We consider that for each $t \in T$ we have a second-order stationary and isotropic random process, that is, the mean and variance functions are constant and the covariance depends only on the distance among sampling points. We want to predict a single variable at a single location from a sample of spatially correlated functional data $\chi_{s_1}(t), \dots, \chi_{s_n}(t)$. Let $\chi_{s_0}(v)$ be the random variable to be predicted at an unsampled location s_0 at a particular temporal instant $v \in T$. To work with cokriging based on curves, we extend the cokriging predictor shown in Section 2 replacing in (1) both the $n \times m$ parameters λ_{ij}^k by n functional parameters $\lambda_i^v(t)$ and the $n \times m$ random variables $Z_j(s_i)$ by n functional variables $\chi_{s_i}(t)$, $i = 1, \dots, n, j = 1, \dots, m$. We describe these replacements in (4) and (5), respectively:

Parameters

$$\begin{array}{ccc}
 \textit{Multivariable Cokriging} & & \textit{Cokriging Based on Curves} \\
 \lambda_{11} \cdots \lambda_{1m} & \Rightarrow & \lambda_1^v(t), t \in T \\
 \lambda_{21} \cdots \lambda_{2m} & \Rightarrow & \lambda_2^v(t), t \in T \\
 \vdots & \vdots & \vdots \\
 \lambda_{n1} \cdots \lambda_{nm} & \Rightarrow & \lambda_n^v(t), t \in T
 \end{array} \quad (4)$$

Variables

$$\begin{array}{ccc}
 \textit{Multivariable Cokriging} & & \textit{Cokriging Based on Curves} \\
 Z_1(s_1) \cdots Z_m(s_1) & \Rightarrow & \chi_{s_1}(t), t \in T \\
 Z_1(s_2) \cdots Z_m(s_2) & \Rightarrow & \chi_{s_2}(t), t \in T \\
 \vdots & \vdots & \vdots \\
 Z_1(s_n) \cdots Z_m(s_n) & \Rightarrow & \chi_{s_n}(t), t \in T
 \end{array} \quad (5)$$

Thus, the cokriging predictor of $\chi_{s_0}(v)$ based on functional data (CBFD) is given by

$$\hat{\chi}_{s_0}(v) = \sum_{i=1}^n \int_T \lambda_i^v(t) \chi_{s_i}(t) dt. \quad (6)$$

For each specific $v \in T$, the functional parameters $\lambda_i^v(t)$, $i = 1, \dots, n$ in (6) are estimated taking into account classical geostatistical constraints, that is, unbiasedness and minimum prediction variance. We solve this problem using an approach based on basis functions. We expand functional variables and parameters by

$$\chi_{s_i}(t) = \sum_{l=1}^K a_{il} B_l(t) = \mathbf{a}_i^T \mathbf{B}(t), \quad (7)$$

and

$$\lambda_i^v(t) = \sum_{l=1}^K b_{ilv} B_l(t) = \mathbf{b}_{iv}^T \mathbf{B}(t), \quad (8)$$

where K defines the number of basis functions, and its value can be obtained using a cross-validation procedure as explained in Section 5. Therefore, using (7) and (8) the predictor (6) is expressed as

$$\begin{aligned} \hat{\chi}_{s_0}(v) &= \sum_{i=1}^n \int_T \mathbf{b}_{iv}^T \mathbf{B}(t) \mathbf{B}^T(t) \mathbf{a}_i dt \\ &= \sum_{i=1}^n \mathbf{b}_{iv}^T \mathbf{W} \mathbf{a}_i, \end{aligned} \quad (9)$$

where

$$\mathbf{W} = \int_T \mathbf{B}(t) \mathbf{B}^T(t) dt. \quad (10)$$

For any orthonormal basis such as the Fourier basis, the Gram matrix \mathbf{W} is the identity matrix. For other basis functions such as B-Splines, \mathbf{W} must be calculated by numerical integration.

Assuming stationarity of the random functions, the matrix

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1K} \\ a_{21} & a_{22} & \cdots & a_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nK} \end{pmatrix} = (\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_K)_{(n \times K)} \quad (11)$$

forms a K multivariable random field with

$$\mathbb{E}(\boldsymbol{\alpha}_j) = \begin{pmatrix} \mathbb{E}(a_{1j}) \\ \mathbb{E}(a_{2j}) \\ \vdots \\ \mathbb{E}(a_{nj}) \end{pmatrix} = \begin{pmatrix} \vartheta_j \\ \vartheta_j \\ \vdots \\ \vartheta_j \end{pmatrix}, j = 1, \dots, K, \quad (12)$$

and covariance matrix

$$\boldsymbol{\Sigma} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} & \cdots & \Sigma_{1K} \\ \Sigma_{21} & \Sigma_{22} & \cdots & \Sigma_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma_{K1} & \Sigma_{K2} & \cdots & \Sigma_{KK} \end{pmatrix}. \quad (13)$$

where $\Sigma_{ij} = \mathbb{C}(\boldsymbol{\alpha}_i, \boldsymbol{\alpha}_j)_{(n \times n)}$, is the covariance between $\boldsymbol{\alpha}_i$ and $\boldsymbol{\alpha}_j$. From equation (12) We have

$$\mathbb{E}(\mathbf{a}_i) = \begin{pmatrix} \mathbb{E}(a_{i1}) \\ \mathbb{E}(a_{i2}) \\ \vdots \\ \mathbb{E}(a_{iK}) \end{pmatrix} = \begin{pmatrix} \vartheta_1 \\ \vartheta_2 \\ \vdots \\ \vartheta_K \end{pmatrix} = \boldsymbol{\vartheta}. \quad (14)$$

Thus, the mean of the predictor (9) is given by

$$\begin{aligned} \mathbb{E}(\hat{\chi}_{s_0}(v)) &= \sum_{i=1}^n \mathbf{b}_{iv}^T \mathbf{W} \mathbb{E}(\mathbf{a}_i) \\ &= \sum_{i=1}^n \mathbf{b}_{iv}^T \mathbf{W} \boldsymbol{\vartheta}. \end{aligned}$$

On the other hand the mean of the unobserved function on site s_0 at a time v is

$$\begin{aligned} \mathbb{E}(\chi_{s_0}(v)) &= \sum_{l=1}^K \mathbb{E}(a_{0l}) B_l(v) \\ &= \mathbf{B}^T(v) \mathbb{E}(\mathbf{a}_0) \\ &= \mathbf{B}^T(v) \boldsymbol{\vartheta}. \end{aligned}$$

Consequently, the proposed predictor is unbiased if

$$\sum_{i=1}^n \mathbf{b}_{iv}^T = \mathbf{B}^T(v) \mathbf{W}^{-1}. \quad (15)$$

Although a Gram matrix is in general positive semidefinite, and positive definiteness can not be generally guaranteed, our Gram matrix \mathbf{W} in (10) is positive definite because the functions $B_l(t), l = 1, \dots, K$ are part of a basis and thus are linearly independent. Consequently, \mathbf{W}^{-1} exists and the equation (15) is well defined.

To find the best linear unbiased predictor (BLUP), the n functional parameters in the proposed predictor are given by the solution of the following optimization problem

$$\underset{\lambda_1^v(\cdot), \dots, \lambda_n^v(\cdot)}{\text{Min}} \quad \mathbb{V}(\hat{\chi}_{s_0}(v) - \chi_{s_0}(v)), \text{ s.t. } \mathbb{E}(\hat{\chi}_{s_0}(v)) = \mathbb{E}(\chi_{s_0}(v)). \quad (16)$$

Developing the variance in the objective function (16) we find the following expression:

$$\begin{aligned}
\mathbb{V}(\hat{\chi}_{s_0}(v) - \chi_{s_0}(v)) &= \mathbb{V}(\hat{\chi}_{s_0}(v)) + \mathbb{V}(\chi_{s_0}(v)) - 2\mathbb{C}(\hat{\chi}_{s_0}(v), \chi_{s_0}(v)) \\
&= \sum_{i=1}^n \mathbf{b}_{iv}^T \mathbf{W} \mathbb{V}(\mathbf{a}_i) \mathbf{W}^T \mathbf{b}_{iv} \\
&\quad + 2 \sum_{i < j} \mathbf{b}_{iv}^T \mathbf{W} \mathbb{C}(\mathbf{a}_i, \mathbf{a}_j) \mathbf{W}^T \mathbf{b}_{jv} \\
&\quad + \mathbf{B}^T(v) \mathbb{V}(\mathbf{a}_0) \mathbf{B}(v) \\
&\quad - 2 \sum_{i=1}^n \mathbf{b}_{iv}^T \mathbf{W} \mathbb{C}(\mathbf{a}_0, \mathbf{a}_i) \mathbf{B}(v), \tag{17}
\end{aligned}$$

where $\mathbb{V}(\mathbf{a}_i)$ and $\mathbb{V}(\mathbf{a}_0)$ stand for the variance of the vectors \mathbf{a}_i and \mathbf{a}_0 respectively, and $\mathbb{C}(\mathbf{a}_i, \mathbf{a}_j)$ and $\mathbb{C}(\mathbf{a}_0, \mathbf{a}_i)$ define, respectively, the covariance between the vectors \mathbf{a}_i and \mathbf{a}_j , and between the vectors \mathbf{a}_0 and \mathbf{a}_i . These values can be calculated if $\boldsymbol{\Sigma}$ in (13) has been previously estimated, but I can use multivariable geostatistics (Wackernagel 1995) and in particular a linear model of coregionalization (LMC) (Borgault & Marcotte 1991) in order to estimate these matrices. Note that because of the stationarity of the random functions in \mathbf{A} , the variances of \mathbf{a}_i and \mathbf{a}_0 are equal. In addition, $\mathbb{C}(\mathbf{a}_0, \mathbf{a}_i)$ depends only on the separation vector and not on the locations themselves. Thus they can be calculated directly from $\boldsymbol{\Sigma}$.

If we define

$$\begin{aligned}
\mathbf{M}_i &= \left(\mathbf{W} \mathbb{V}(\mathbf{a}_i) \mathbf{W}^T \right)_{(K \times K)}, \\
\mathbf{M}_{ij} &= \left(\mathbf{W} \mathbb{C}(\mathbf{a}_i, \mathbf{a}_j) \mathbf{W}^T \right)_{(K \times K)} \\
\mathbf{N}_i(v) &= \left(\mathbf{W} \mathbb{C}(\mathbf{a}_0, \mathbf{a}_i) \mathbf{B}(v) \right)_{(K \times 1)} \tag{18}
\end{aligned}$$

and

$$\mathbf{D}(v) = \mathbf{B}^T(v) \mathbb{V}(\mathbf{a}_0) \mathbf{B}(v), \tag{19}$$

equation (17) can be rewritten as:

$$\begin{aligned}
\mathbb{V}(\hat{\chi}_{s_0}(v) - \chi_{s_0}(v)) &= \sum_{i=1}^n \mathbf{b}_{iv}^T \mathbf{M}_i \mathbf{b}_{iv} + 2 \sum_{i < j} \mathbf{b}_{iv}^T \mathbf{M}_{ij} \mathbf{b}_{jv} \\
&\quad + \mathbf{D}(v) - 2 \sum_{i=1}^n \mathbf{b}_{iv}^T \mathbf{N}_i(v). \tag{20}
\end{aligned}$$

From (20) and considering K Lagrange multipliers $\mathbf{m}_v^T = (m_{1v}, \dots, m_{Kv})$, the

optimization problem (16) can be expressed as

$$\min_{\mathbf{b}_{1v}, \dots, \mathbf{b}_{nv}, \mathbf{m}_v} \sum_{i=1}^n \mathbf{b}_{iv}^T \mathbf{M}_i \mathbf{b}_{iv} + 2 \sum_{i < j} \mathbf{b}_{iv}^T \mathbf{M}_{ij} \mathbf{b}_{jv} + \mathbf{D}(v) - 2 \sum_{i=1}^n \mathbf{b}_{iv}^T \mathbf{N}_i(v) + 2 \mathbf{m}_v^T \left(\sum_{i=1}^n \mathbf{b}_{iv} - \mathbf{W}^{-1} \mathbf{B}(v) \right). \quad (21)$$

Taking $\boldsymbol{\beta}_v = (\mathbf{b}_{1v}^T, \dots, \mathbf{b}_{nv}^T, \mathbf{m}_v^T)^T_{(K(n+1) \times 1)}$, then is given by

$$\min_{\boldsymbol{\beta}_v} \boldsymbol{\beta}_v^T \mathbf{M} \boldsymbol{\beta}_v + \mathbf{D}(v) - 2 \boldsymbol{\beta}_v^T \mathbf{N}(v) \quad (22)$$

where

$$\mathbf{M} = \begin{pmatrix} \mathbf{M}_1 & \mathbf{M}_{12} & \cdots & \mathbf{M}_{1n} & \mathbf{I} \\ \mathbf{M}_{21} & \mathbf{M}_2 & \cdots & \mathbf{M}_{2n} & \mathbf{I} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{M}_{n1} & \mathbf{M}_{n2} & \cdots & \mathbf{M}_n & \mathbf{I} \\ \mathbf{I} & \mathbf{I} & \cdots & \mathbf{I} & \mathbf{0} \end{pmatrix}_{[K(n+1)] \times [K(n+1)]}$$

and

$$\mathbf{N}(v) = \begin{pmatrix} \mathbf{N}_1(v) \\ \mathbf{N}_2(v) \\ \vdots \\ \mathbf{N}_n(v) \\ \mathbf{W}^{-1} \mathbf{B}(v) \end{pmatrix}_{(K(n+1) \times 1)} \quad (23)$$

Finally, minimizing equation (22) with respect to $\boldsymbol{\beta}_v$ I obtain

$$2\mathbf{M}\boldsymbol{\beta}_v - 2\mathbf{N}(v) = \mathbf{0} \Rightarrow \mathbf{M}\boldsymbol{\beta}_v = \mathbf{N}(v) \Rightarrow \hat{\boldsymbol{\beta}}_v = \mathbf{M}^{-1}\mathbf{N}(v). \quad (24)$$

A plug-in estimation of the prediction variance $\sigma_{s_0}^2(v) = \mathbb{V}(\hat{\chi}_{s_0}(v) - \chi_{s_0}(v))$ is given by

$$\hat{\sigma}_{s_0}^2(v) = \hat{\boldsymbol{\beta}}_v^T \mathbf{M} \hat{\boldsymbol{\beta}}_v + \mathbf{D}(v) - 2 \hat{\boldsymbol{\beta}}_v^T \mathbf{N}(v), \quad (25)$$

where the matrix $\mathbf{D}(v)$ defined as in equation (19) is calculated using an estimate of $\hat{\mathbb{V}}(\mathbf{a}_0)$ obtained by means of the fitted LMC.

4 Functional kriging: total model

In this section we show a predictor defined in Giraldo (2009) which is also considered by Nerini et al. (2010). We call this one functional kriging (*total model*)

predictor (FKTM). We assume the same stationarity and isotropy assumptions as for the cokriging in Section 3. The cokriging predictor given in (6) is defined for an specific $v \in T$. If we want to predict the whole curve at s_0 , the functional parameter $\lambda_i^v(t)$ in (6) is replaced by a double indexed functional parameter $\lambda_i(t, v)$. Thus, the predictor of the whole curve is given by

$$\hat{\chi}_{s_0}(v) = \sum_{i=1}^n \int_T \lambda_i(t, v) \chi_{s_i}(t) dt, \quad v \in T, \quad (26)$$

such that $\lambda_1(t, v), \dots, \lambda_n(t, v) : T \times T \rightarrow \mathbb{R}$. Note that, following the replacements given in the schemes (4) and (5), the predictor given in (26) is an extension to the functional context of the predictor (3). It is clear that, for a fixed $v \in T$, the FKTM predictor in equation (26) and the Cbfd predictor given in (6) have similar expressions. Proposition 1 below establishes the equivalence between both problems when basis function expansion is used to solve them.

The functional parameter $\lambda_i(t, v)$ in (26) determines the impact of the i -th observed function at time t on an unobserved function at time v . This modeling approach is coherent with the functional linear model for functional responses (*total model*) shown in Ramsay & Silverman (2005). In that framework and assuming that we generally have $X_j \in L_2(T_1), Y \in L_2(T_2)$ with $j = 1, \dots, q$, with L_2 a space of square integrable functions, the functional response $Y(v)$ is modeled in terms of the functional covariates by

$$Y(v) = \alpha(v) + \sum_{j=1}^q \int_{T_1} X_j(t) \beta_j(t, v) dt + \epsilon(v), \quad (27)$$

where $\beta_j \in L_2(T_1 \times T_2)$ is a parameter function, $\alpha \in L_2(T_2)$ is an intercept function and $\epsilon \in L_2(T_2)$ is a random error process such that $\mathbb{E}(\epsilon(v)) = 0$ for all v . Estimation of functional parameters in (27) is carried out by solving (Ramsay & Silverman 2005)

$$\underset{\alpha(\cdot), \beta_1(\cdot, \cdot), \dots, \beta_q(\cdot, \cdot)}{\text{Min}} \mathbb{E} \|\hat{Y}(\cdot) - Y(\cdot)\|^2.$$

In our context the covariates are the observed curves in n sites of a region and the functional response is an unobserved function at an unvisited location. Consequently our objective function, based on the L_2 -norm is

$$\mathbb{E} \|\hat{\chi}_{s_0}(\cdot) - \chi_{s_0}(\cdot)\|_2^2,$$

depending on $\lambda_1(\cdot, \cdot), \dots, \lambda_n(\cdot, \cdot)$, or by using Fubini's Theorem

$$\int_T \mathbb{E} (\hat{\chi}_{s_0}(v) - \chi_{s_0}(v))^2 dv.$$

Considering stationarity, the objective function becomes

$$\int_T \mathbb{V} (\hat{\chi}_{s_0}(v) - \chi_{s_0}(v)) dv.$$

Again the functional parameters $\lambda_i(t, v)$ in (26) are estimated taking into account the constraints of unbiasedness and minimum prediction variance. Thus, the optimization problem becomes

$$\underset{\lambda_1(\cdot, \cdot), \dots, \lambda_n(\cdot, \cdot)}{\text{Min}} \int_T \mathbb{V}(\hat{\chi}_{s_0}(v) - \chi_{s_0}(v)) dv \text{ s.t. } \mathbb{E}(\hat{\chi}_{s_0}(v)) = \mathbb{E}(\chi_{s_0}(v)), \forall v \in T.$$

We solve this problem by using an approach based on basis functions. We expand the functional variables as in the equation (7) and the bivariate functional parameters by

$$\lambda_i(t, v) = \sum_{j=1}^K \sum_{l=1}^K c_{jl}^i B_j(t) B_l(v) = \mathbf{B}^T(t) \mathbf{C}_i \mathbf{B}(v), \quad (28)$$

where

$$\mathbf{C}_i = \begin{pmatrix} c_{11}^i & c_{12}^i & \cdots & c_{1K}^i \\ c_{21}^i & c_{22}^i & \cdots & c_{2K}^i \\ \vdots & \vdots & \ddots & \vdots \\ c_{K1}^i & c_{K2}^i & \cdots & c_{KK}^i \end{pmatrix}_{(K \times K)}.$$

From equations (7) and (28), the predictor (26) can be expressed as

$$\begin{aligned} \hat{\chi}_{s_0}(v) &= \sum_{i=1}^n \int_T \mathbf{a}_i^T \mathbf{B}(t) \mathbf{B}^T(t) \mathbf{C}_i \mathbf{B}(v) dt \\ &= \sum_{i=1}^n \mathbf{a}_i^T \mathbf{W} \mathbf{C}_i \mathbf{B}(v) = \mathbf{B}^T(v) \sum_{i=1}^n \mathbf{C}_i^T \mathbf{W}^T \mathbf{a}_i \\ &= \mathbf{B}^T(v) \sum_{i=1}^n \mathbf{C}_i^T \mathbf{W} \mathbf{a}_i = \mathbf{B}^T(v) \hat{\mathbf{a}}_0, \end{aligned} \quad (29)$$

where the matrix of inner products \mathbf{W} is defined as in equation (10). The predictor (26) is also considered by Nerini et al. (2010). These authors assume that \mathbf{W} is the identity matrix because they consider a solution based on orthonormal basis expansions. In our solution this is not a necessary condition. Now we consider the unbiasedness and minimum variance properties of the proposed predictor. As in Section 3, we assume that the coefficients \mathbf{a}_i in the equation (29) follow a stationary multivariable random field. Consequently from equation (14) the expected value of the curve on an unvisited site s_0 is given by

$$\begin{aligned} \mathbb{E}(\chi_{s_0}(v)) &= \mathbb{E} \left(\sum_{j=1}^K a_{0j} B_j(v) \right) \\ &= \mathbb{E}(\mathbf{B}^T(v) \mathbf{a}_0) = \mathbf{B}^T(v) \mathbb{E}(\mathbf{a}_0) \\ &= \mathbf{B}^T(v) \boldsymbol{\vartheta}. \end{aligned} \quad (30)$$

On the other hand taking expected values in (29) we have

$$\begin{aligned}\mathbb{E}(\hat{\chi}_{s_0}(v)) &= \mathbf{B}^T(v) \sum_{i=1}^n \mathbf{C}_i^T \mathbf{W} \mathbb{E}(\mathbf{a}_i) \\ &= \mathbf{B}^T(v) \sum_{i=1}^n \mathbf{C}_i^T \mathbf{W} \boldsymbol{\vartheta}.\end{aligned}\quad (31)$$

Consequently from equations (30) and (31) we note that the predictor (26) is unbiased if and only if

$$\mathbf{B}^T(v) \sum_{i=1}^n \mathbf{C}_i^T \mathbf{W} \boldsymbol{\vartheta} = \mathbf{B}^T(v) \boldsymbol{\vartheta}, \text{ for all } \boldsymbol{\vartheta} \in T,$$

that is, if and only if,

$$\sum_{i=1}^n \mathbf{C}_i^T \mathbf{W} \boldsymbol{\vartheta} = \boldsymbol{\vartheta}, \text{ for all } \boldsymbol{\vartheta} \in T.$$

Given that W is full rank, this condition is equivalent to

$$\sum_{i=1}^n \mathbf{C}_i = \mathbf{W}^{-1}.$$

The n functional parameters in the predictor (26) are given by the solution of the following optimization problem

$$\underset{\mathbf{C}_1, \dots, \mathbf{C}_n}{Min} \int_T \mathbb{V}(\mathbf{B}^T(v) \hat{\mathbf{a}}_0 - \mathbf{B}^T(v) \mathbf{a}_0) dv \text{ s.t. } \sum_{i=1}^n \mathbf{C}_i = \mathbf{W}^{-1}. \quad (32)$$

The integral in the objective function (32) can be rewritten as

$$\begin{aligned}& \int_T \mathbf{B}^T(v) \mathbb{V}(\hat{\mathbf{a}}_0 - \mathbf{a}_0) \mathbf{B}(v) dv \\ &= \int_T Tr \left(\mathbf{B}^T(v) \mathbb{V}(\hat{\mathbf{a}}_0 - \mathbf{a}_0) \mathbf{B}(v) \right) dv \\ &= Tr \left(\mathbb{V}(\hat{\mathbf{a}}_0 - \mathbf{a}_0) \int_T \mathbf{B}(v) \mathbf{B}^T(v) dv \right) \\ &= Tr(\mathbb{V}(\hat{\mathbf{a}}_0 - \mathbf{a}_0) \mathbf{W}).\end{aligned}\quad (33)$$

The variance in (33) is

$$\begin{aligned}\mathbb{V}(\hat{\mathbf{a}}_0 - \mathbf{a}_0) &= \sum_{i=1}^n \mathbf{C}_i^T \mathbf{W} \mathbb{V}(\mathbf{a}_i) \mathbf{W} \mathbf{C}_i + 2 \sum_{i < j}^n \mathbf{C}_i^T \mathbf{W} \mathbb{C}(\mathbf{a}_i, \mathbf{a}_j) \mathbf{W} \mathbf{C}_j \\ &\quad + \mathbb{V}(\mathbf{a}_0) - 2 \sum_{i=1}^n \mathbf{C}_i^T \mathbf{W} \mathbb{C}(\mathbf{a}_i, \mathbf{a}_0).\end{aligned}\quad (34)$$

From (34) and defining the following matrices

$$\mathbf{Q}_{ii} = \mathbf{W}\mathbb{V}(\mathbf{a}_i)\mathbf{W}, \mathbf{Q}_{ij} = \mathbf{W}\mathbb{C}(\mathbf{a}_i, \mathbf{a}_j)\mathbf{W}, \mathbf{N}_i = \mathbf{W}\mathbb{C}(\mathbf{a}_i, \mathbf{a}_0)$$

the optimization problem (32) can be expressed as

$$\begin{aligned} \underset{\mathbf{C}_1, \dots, \mathbf{C}_n, \mathbf{m}}{\text{Min}} \quad & \sum_{i=1}^n \text{Tr} \left(\mathbf{C}_i^T \mathbf{Q}_{ii} \mathbf{C}_i \mathbf{W} \right) + 2 \sum_{i < j}^n \text{Tr} \left(\mathbf{C}_i^T \mathbf{Q}_{ij} \mathbf{C}_j \mathbf{W} \right) + \text{Tr}(\mathbb{V}(\mathbf{a}_0)\mathbf{W}) \\ & - 2 \sum_{i=1}^n \text{Tr} \left(\mathbf{C}_i^T \mathbf{N}_i \mathbf{W} \right) + 2\mathbf{m} \left(\sum_{i=1}^n \mathbf{C}_i - \mathbf{W}^{-1} \right). \end{aligned} \quad (35)$$

Derivatives with respect to $\mathbf{C}_i, i = 1, \dots, n$, and \mathbf{m} in (35) are given, respectively, by

$$2 \sum_{j=1}^n \mathbf{Q}_{ij} \mathbf{C}_j \mathbf{W} - 2\mathbf{N}_i \mathbf{W} + 2\mathbf{m}, \quad \text{and} \quad \sum_{i=1}^n \mathbf{C}_i - \mathbf{W}^{-1}.$$

The solution of the problem given in (35) is achieved by setting these derivatives equal to zero. This solution can be represented in matrix notation as

$$\begin{pmatrix} \mathbf{Q}_{11} & \mathbf{Q}_{12} & \cdots & \mathbf{Q}_{1n} & \mathbf{I} \\ \mathbf{Q}_{21} & \mathbf{Q}_{22} & \cdots & \mathbf{Q}_{2n} & \mathbf{I} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{Q}_{n1} & \mathbf{Q}_{n2} & \cdots & \mathbf{Q}_{nn} & \mathbf{I} \\ \mathbf{I} & \mathbf{I} & \cdots & \mathbf{I} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{C}_1 \\ \mathbf{C}_2 \\ \vdots \\ \mathbf{C}_n \\ \mathbf{m}^* \end{pmatrix} = \begin{pmatrix} \mathbf{N}_1 \\ \mathbf{N}_2 \\ \vdots \\ \mathbf{N}_n \\ \mathbf{W}^{-1} \end{pmatrix}, \quad (36)$$

where $\mathbf{m}^* = \mathbf{m}\mathbf{W}^{-1}$. From equations (33) and (34), an estimation of the integrated prediction variance

$$\sigma_{int}^2 = \int_T \sigma_{s_0}^2(v) dv = \int_T \mathbb{V}(\hat{\chi}_{s_0}(v) - \chi_{s_0}(v)) dv$$

is given by

$$\begin{aligned} \hat{\sigma}_{int}^2 = & \sum_{i=1}^n \text{Tr} \left(\hat{\mathbf{C}}_i^T \mathbf{Q}_{ii} \hat{\mathbf{C}}_i \mathbf{W} \right) + 2 \sum_{i < j}^n \text{Tr} \left(\hat{\mathbf{C}}_i^T \mathbf{Q}_{ij} \hat{\mathbf{C}}_j \mathbf{W} \right) \\ & + \text{Tr}(\mathbb{V}(\mathbf{a}_0)\mathbf{W}) - 2 \sum_{i=1}^n \text{Tr} \left(\hat{\mathbf{C}}_i \mathbf{N}_i \mathbf{W} \right), \end{aligned} \quad (37)$$

where the matrices $\hat{\mathbf{C}}_1, \dots, \hat{\mathbf{C}}_n$ are obtained by solving the system (36).

To finish with this section we note the connection between the two methods introduced in the paper. The relationship between CBFD and FKTM is analogous to that of cokriging analysis and multivariable spatial prediction, in the sense that the prediction obtained by CBFD at a time is identical to the prediction achieved by FKTM at the same time.

The expressions of these predictors, the unbiasedness constraints and the respective objective functions, are equivalent at each fixed time v . The distinctive feature between these approaches is given in terms of the prediction variances. In the first case a prediction variance at each particular time can be obtained, whereas in the second one an integrated prediction variance can be estimated and used as a *global* measure of uncertainty. Indeed Proposition 1 establishes the equivalence between both alternatives when, as we propose, a basis function expansion is used. The proof is deferred to the Appendix.

Proposition 1. *Assume that representation (7) is correct for χ_{s_i} , $i = 0, 1, \dots, n$. Let $\hat{\chi}_{s_0}^C(v)$, $v \in T$, be the set of predictors for $\chi_{s_0}(v)$, $v \in T$, derived from equations (6) and (8), where the coefficients \mathbf{b}_{iv} , $i = 0, 1, \dots, n$, are the solution of problem (22) for each $v \in T$. Let $\hat{\chi}_{s_0}^F(v)$, $v \in T$, be the predicted function given by equations (26) and (28), where coefficient matrices \mathbf{C}_i , $i = 0, 1, \dots, n$, are the solution of problem (35). Then, assuming that we use the L_2 -norm, i.e., the norm of χ_{s_i} is the expectation of the integral of squared $\chi_{s_i}(v)$ with $v \in T$,*

$$\hat{\chi}_{s_0}^C(v) = \hat{\chi}_{s_0}^F(v)$$

for all $v \in T$ and all $s_0 \in D$.

5 Applications

In applied sciences, it is common that data have both spatial and functional components. One such example could be meteorology, when curves of temperature or precipitation are obtained in several weather stations of a country (Ramsay & Silverman 2005). Likewise in agronomy when measures of penetration resistance are taken in a sampling grid of the study area (Chan et al. 2006). In this case, and though penetration resistance is measured only at some depths, it is possible to consider it as a functional variable after a smoothing or interpolation process have been applied. We consider two real data sets in these fields to illustrate the approach.

5.1 Spatial prediction of temperature curves in the Canadian Maritime Provinces

Spatial prediction of meteorological data is an important input for many types of models. In particular, the modeling of spatially correlated temperature data is of interest, among others, for predicting microclimate conditions in mountainous terrain, resource management, calibration of satellite sensors or for studying the “greenhouse effect”.

A data set frequently used in the FDA setting corresponds to temperature curves recorded at 35 weather stations located throughout Canada (Yamanishi & Tanaka (2003), Ramsay & Silverman (2005) and Hall & Hosseini-Nassab (2006)). This

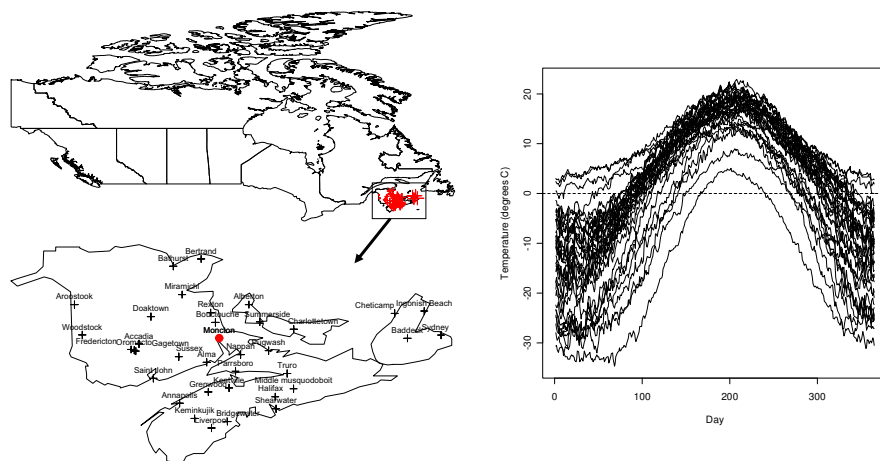


Figure 1: *Averages (over 30 years) of daily temperature curves (right panel) observed at 35 weather stations of the Canadian Maritime provinces (left panel). Source: own elaboration.*

corresponds to a spatially correlated functional data set showing an unusual large spatial scale as the weather stations expand over the Canadian territory. The large distances among the locations are reflected in marked differences among the temperatures values, making the assumptions of stationarity and isotropy hard to meet. Taking into account that we assume stationarity and isotropy in the model introduced here, we analyze a homogeneous smaller area in Canada (the Maritime Provinces), thereby assuming negligible the above-mentioned effects. In particular we use a data set consisting of temperature measurements recorded at 35 weather stations located in the Canadian Maritime Provinces (Figure 1, left panel). The Maritime Provinces cover a region of Canada consisting of three provinces: Nova Scotia (NS), New Brunswick (NB), and Prince Edward Island (PEI). The temperature here is characterized by cool summers and mild winters, with a much smaller annual temperature range than that recorded in other Canadian regions. The effects of longitude, latitude, coastal and inland climates are unimportant, and the conditions of stationarity and isotropy are therefore plausible (Stanley 2002).

We thus analyze information of daily temperatures averaged over the years 1960 to 1994 (February 29th combined with February 28th) (Figure 1, right panel). The mean of the averaged daily temperatures varies from -9.4°C in wintertime to 19.3°C in summer. The data for each station were obtained from the Meteorological Service of Canada (<http://www.climate.weatheroffice.ec.gc.ca/climateData/>). The geographical coordinates in decimal degrees of the weather stations (Figure 1, left panel) were obtained from the database of geographic coordinate information (<http://www.tageo.com>).

When data are periodic, Fourier basis with an even number of basis functions is the most appropriate choice (Ramsay & Silverman 2005). To choose an appropriate

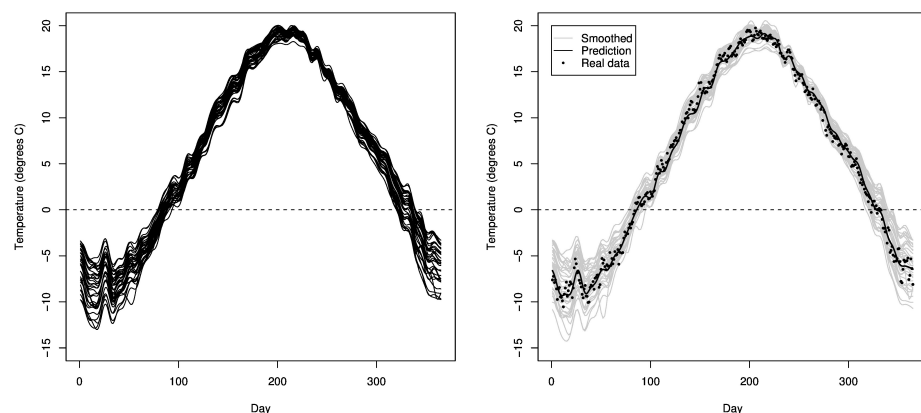


Figure 2: Left panel: Cross-validation predictions obtained by FKTM. Right panel: Predicted temperature at Moncton station obtained by FKTM (dark line), smoothed temperature curves (gray color), and real data recorded at Moncton (circles). Source: own elaboration.

number of basis functions to smooth the discrete data set recorded at each weather station, we followed a non-parametric cross-validation criterion (Giraldo et al. 2010). This criterion is defined in terms of the sum of squared errors (SSE) of cross-validation by

$$\text{NPCV}(K) = \sum_{i=1}^n \|\tilde{\chi}_{s_i}(t) - \chi_{s_i}(t)\| = \sum_{i=1}^n \sum_{j=1}^M (\tilde{\chi}_{s_i}^{(j)}(t_j) - \chi_{s_i}(t_j))^2,$$

where $\tilde{\chi}_{s_i}^{(j)}(t_j)$ is the estimated function at t_j by means of equation (9) when the datum $\chi_{s_i}(t_j)$ has been temporarily suppressed from the sample. The strategy is to minimize $\text{NPCV}(K)$ in K . We concluded that a basis with 65 functions is appropriate for smoothing the temperature values recorded in the Canadian Maritime provinces data set. This number agrees with that reported by Giraldo et al. (2010) and Ramsay & Silverman (2005).

To verify the goodness-of-fit of the proposed predictors and to compare them with other spatial predictors for functional data, we used a functional cross-validation analysis. Each individual smoothed curve $\chi_{s_i}(t)$ for $i = 1, \dots, 35$, was temporarily removed, and further predicted from the remaining ones. We also do prediction at an unvisited site (the Moncton station).

Considering the result in Proposition 1, we know that predictions obtained by CBFD coincide with those given by FKTM. We thus only show the results obtained by FKTM. Once the coefficients in matrix (11) were estimated, a LMC was fitted to empirical variograms and cross-variograms and used to calculate the matrices \mathbf{Q}_{ij} , \mathbf{N}_i , $i, j = 1, \dots, 35$, in equation (36), to finally estimate the parameters given

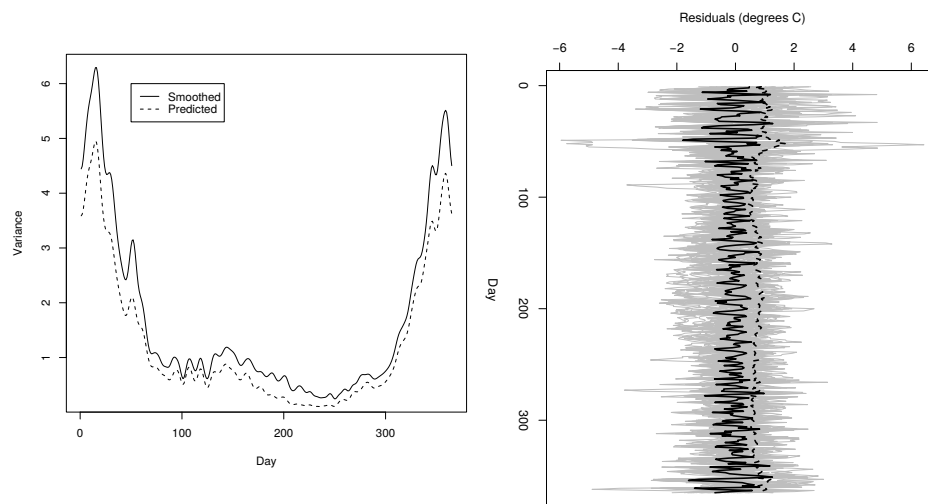


Figure 3: *Left panel: Pointwise variance calculated with the smoothed curves and the predicted curves (obtained by FKTM). Right panel: Cross-validation residuals obtained by FKTM (clear lines), residual mean (dark line) and residual standard deviation (dashed line). Source: own elaboration.*

in that system. According to the LMC, we assumed that all direct and cross-variograms should be modeled as linear combinations of the same set of variogram models (nugget, spherical, exponential, Gaussian, etc). The results coming from several combinations of tentative models were compared graphically, and then all direct and cross-variograms were modeled as a linear combination of nugget and exponential models. In this step of the analysis we used the library `gstat` of the R language (Pebesma 2004).

Figure 2 shows the cross-validation predictions, the smoothed curves (when using 65 Fourier basis functions), and the prediction at Moncton (an unvisited weather station, see Figure 1). A graphical comparison between smoothed (right panel of Figure 2) and predicted curves (left panel of Figure 2) shows the good performance of the predictions. In addition, the kriging prediction at Moncton station is certainly close to the original data (right panel of Figure 2). For each t we calculate the temperature variance based on the smoothed and predicted data (Figure 3, left panel). Note that for each case (data smoothed or predicted) these variances are calculated with 35 data (one data for each weather station). The predicted curves show less pointwise variance than the smoothed ones which is not surprising, because kriging is itself a smoothing method.

Right panel of Figure 3 shows the cross-validation residuals (differences between the smoothed curves and the predicted curves). In general, good predictions in a high proportion of sites (those having residuals around zero) are obtained. Large positive or negative residuals are obtained in just a few number of stations. In

wintertime, the residuals are bigger than in other seasons (residuals higher than 4 °C and lower than -4 °C at the beginning and end of the year), that is, we have greater uncertainty on the prediction in this period. This is due to the fact that the observed temperature curves show more variability at this time of year (Figure 3, left panel). The residual standard deviation is lower in the summer (Figure 3, right panel) where the smoothed and the predicted curves have less variation (Figure 3, left panel). The residual mean varies around zero indicating that the predictions are unbiased (Figure 3, right panel).

Bertrand and Bathurst in the north eastern area of New Brunswick (Figure 1) are the stations with the greatest cross-validation residuals. Though the average daily temperature at these stations has a very similar behavior throughout the time, the difference among its values at some days (19, 49 to 57, 353, 354 and 356) is greater than 4 °C. This fact generates high residuals in both cases, taking into account that the prediction at Bertrand is highly influenced by the curve of Bathurst and vice-versa, because these sites are relatively close (Figure 1).

With respect to the prediction at Moncton we observe that the predicted curve shows a seasonal behavior similar to the smoothed curves (Figure 2, right panel). This is consistent with real values recorded at this weather station. Three of the estimated functional parameters used for prediction at Moncton are shown in Figure 4. These correspond to Bouctouche, Nappan and Aroostook. In Figure 4 we also show the estimated functional parameters for all the stations when predicting the temperature at Moncton on 9th April (day 100). Two relevant aspects can be highlighted from Figure 4: (a) First, Bouctouche and Nappan stations have greater influence in predicting Moncton than Aroostook (whose functional parameter is almost zero). A measure of influence is given by $\|\lambda\|_2$. Numerical values of $\|\lambda\|_2$ for these three stations were 18.7, 3.9, and 0.006, respectively. This result is coherent with the geostatistical philosophy: sites closer to the prediction location have greater influence than others more far apart. Bouctouche and Nappan, separated from Moncton 40 km and 56 km respectively, are the nearest stations to this site in the data set (Figure 1), whereas Aroostook, located in western New Brunswick, is one of the most separated stations (approximately 310 kilometers from Moncton). As Aroostook, other stations far from Moncton have low influence on the prediction; (b) Second, the estimated functional parameters reveal a short temporal effect on the prediction. This was also expected. An analysis of time series, not included in the paper, indicates that these temperatures, considered as time series and differentiated, have low autoregressive orders. This feature is reflected in Figure 4 (bottom, right panel) which corresponds to predicting the temperature at Moncton on 9th April (day 100). The estimated parameters reveal a temporal effect on the prediction even lower than 10 days. The greatest estimated parameter (with values close to 0.09) corresponds to Bouctouche (the nearest station to Moncton) on the same day. Comparing these curves with those corresponding to predicting the temperature at Moncton on a different day, it can be seen that the effect of changing v (the day where the prediction is done) is only reflected at the time where the estimated parameters show their largest values, but not in the shape of the curves.

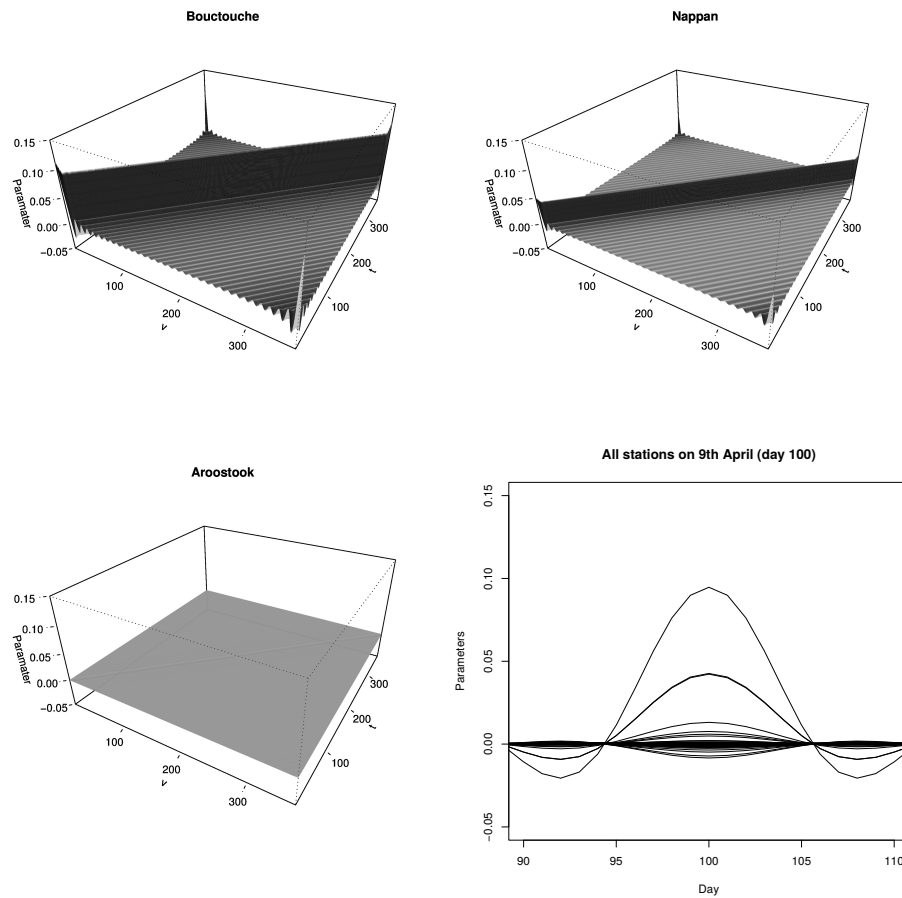


Figure 4: *Top: Estimated functional parameters for Bouctouche (NB) and Nappan (NB). Bottom: Estimated functional parameter for Aroostook (NB) and functional parameters for all stations when we predict at Moncton on 9th April (day 100). Source: own elaboration.*

Table 1: *Summary statistics of the sum of the squared errors of functional cross-validation. Results obtained when using Fourier basis to smooth each curve. OKFD: Ordinary kriging for function-valued spatial data; CTKFD: Continuous time-varying kriging for functional data; FKTM: Functional kriging (total model), that equals CBFD (Cokriging based on functional data). Source: own elaboration.*

Statistic	OKFD	CTKFD	FKTM (or CBFD)
Minimum	103.7	104.4	104.3
Median	253.4	252.7	252.8
Mean	299.5	299.2	298.9
Maximum	890.8	902.1	899.8
Standard deviation	178.4	175.4	176.1
Sum	10484	10471	10461

We have introduced in other previous works some methods for spatial prediction of functional data based on classical kriging approaches. These are called ordinary kriging for function-valued spatial data (OKFD) (Giraldo et al. 2011) and continuous time-varying kriging for spatial prediction of functional data (CTKFD) (Giraldo et al. 2010) which are defined respectively by the following predictors

$$\begin{aligned} \text{OKFD: } \hat{\chi}_{s_0}(t) &= \sum_{i=1}^n \lambda_i \hat{\chi}_{s_i}(t), \quad t \in T \\ \text{CTKFD: } \hat{\chi}_{s_0}(t) &= \sum_{i=1}^n \lambda_i(t) \hat{\chi}_{s_i}(t), \quad t \in T. \end{aligned}$$

The OKFD predictor has the same expression than a classical ordinary kriging but considering curves instead of variables, that is, the predicted curve is a linear combination of observed curves. On the other hand the CTKFD predictor includes functional parameters into the predictor. In this case for each $t \in T$, the predictor equals the ordinary kriging predictor.

We now use the Canadian Maritime Provinces temperature data set and the *functional cross-validation criterion*, as previously described, for comparison purposes. In particular, we evaluate over $j = 1, \dots, 365$ the cross-validation predictions obtained with OKFD, CTKFD, and FKTM (that equals a CBFD predictor when calculating $\hat{\chi}_{s_0}(v)$ in equation (6) for each $v = 1, \dots, 365$). Summary statistics of the sum of the squared errors of *functional cross-validation* (SSE_F) are shown in Table 1. Note that using Fourier basis implies that the matrix W in (10) is an identity matrix, as in the case of (Nerini et al. 2010). In terms of comparisons, and to show the generality of our approach, we additionally carried out an analysis considering B-splines basis (with 65 functions), and thus going far from orthogonality. The results are given in Table 2. A comparison of the SSE_F values by means of a Friedman test (Hollander & Wolfe 1999) shows no significant difference among the six approaches (Friedman chi-squared = 8.0612, df = 5, p-value = 0.1529). Though the temperature data are periodic and consequently a Fourier basis can be more appropriate, the results shown in Table 2 indicate that a B-splines basis could also be useful in this case.

Table 2: *Summary statistics of the sum of the squared errors of functional cross-validation. Results obtained when using B-Splines basis to smooth each curve. OKFD: Ordinary kriging for function-valued spatial data; CTKFD: Continuous time-varying kriging for functional data; FKTM: Functional kriging (total model), that equals CBFD (Cokriging based on functional data). Source: own elaboration.*

Statistic	OKFD	CTKFD	FKTM (or CBFD)
Minimum	106.2	106.3	106.1
Median	251.2	251.2	251.3
Mean	299.5	298.3	298.2
Maximum	899.5	899.5	897.4
Standard deviation	176.4	177.2	176.9
Sum	10483	10441	10437

5.2 Spatial prediction of penetration resistance curves

Penetration resistance is an empirical measure of soil strength that can rapidly identify areas where soil depth or soil compaction may be limiting yields (Chan et al. 2006). The soil mechanical resistance to penetration shows high influence on vegetal development since the growth of the roots and the crops productivity change in an inversely proportional form with its value (Freddi et al. 2006). Knowing its spatial variability provides possibilities for site-specific soil treatments that can increase profitability and sustainability of crop production. Determination of soil resistance at different depths is also necessary to establish proper management strategies. We use here some soil penetration resistance profiles (Figure 5, right panel) obtained on a regular georeferenced grid with 32 sampling points over an agricultural farm at the National University of Colombia (Figure 5, left panel). The study area is situated at $4^{\circ}, 42\text{N}$, $74^{\circ}, 12\text{W}$. For each sampling point, 256 observations of penetration resistance (MPa) were obtained on depths varying from 0 to 35 cm. The positions of the locations were measured by GPS. The data set was obtained by Galindo (2004) as part of a research project focused in precision agriculture (Leiva 2003).

The goal of analyzing this type of data is to predict penetration resistance on unsampled sites based on the collected information, in order to carry out precision agriculture. We note that the curves in Figure 5 are not periodic and consequently a Fourier basis is not the most appropriate choice. Following a non-parametric cross-validation criterion, as shown in previous Section, we used a B-spline basis with 14 basis functions to smooth the data. This was also the number of functions used by Giraldo (2009).

With this basis, the methodology proposed by Nerini et al. (2010) cannot be applied because the matrix W in (10) is not an identity matrix. The set of smoothed curves when using a B-splines basis with 14 functions is shown in Figure 6 (right panel). As an example of the proposed methodology, FKTM prediction on an unvisited location with coordinates 11179 (longitude) and 9750 (latitude) (see left

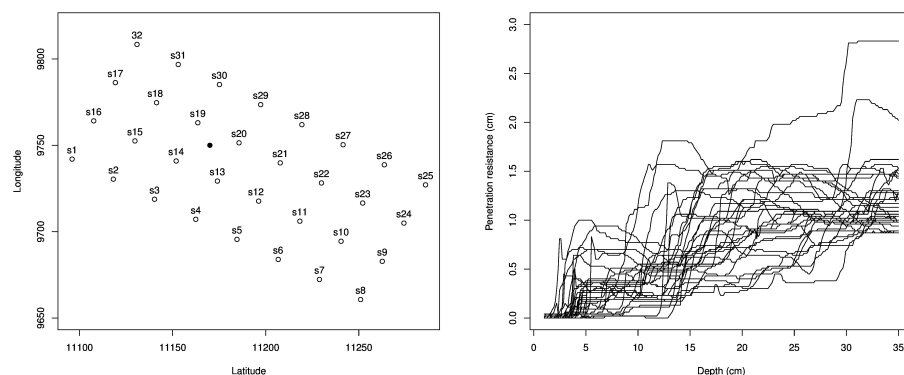


Figure 5: *Observed penetration resistance curves (right panel) on 32 sampling points (left panel). Data collected during 2004 at Marengo Experimental Station (National University of Colombia). Dark point without label in (left panel) corresponds to a unvisited site. Source: own elaboration.*

panel in Figure 5) was performed. As in Section 5.1 a LMC was estimated and used to calculate the matrices \mathbf{Q}_{ij} , \mathbf{N}_i , $i, j = 1, \dots, 32$, in equation (36), and consequently to estimate the parameters given in that system. Any individual variogram and cross-variograms were modeled as a linear combination of nugget and exponential models. The predicted curve (Figure 6, solid line in right panel) indicates that in this location there is a good soil compaction level, because the predicted penetration resistance is less than 2 MPa, which is considered the critical limit for root growth (Chan et al. 2006).

Three of the estimated functional parameters used for the prediction at the unvisited site are shown in Figure 7. These correspond to sites 19, 20 and 8 (see Figure 5). From Figure 7, it is clear that sites 19 and 20 have greater influence than others in the prediction (whose functional parameter is almost zero). A comparison in terms of the $\|\lambda\|_2$ (2.09, 1.68, and 1.9×10^{-6} , respectively) confirms this aspect.

From a geostatistical point of view this is an expected result because sites 19 and 20 are closer to the prediction site than site 8 (Figure 5). As site 8, other stations far from the unvisited site have low influence on the prediction. In Figure 7 we also show the estimated functional parameters for all the 32 sites when predicting the penetration resistance at the unvisited site on depth 5 cm. Here we observe that four sites have greater influence than other in the prediction. These corresponds to sites 19, 20, 13, and 14, that is, the four sites closest to the prediction site (Figure 5).

As in Section 5.1 we carried out a cross-validation analysis with the penetration resistance data set. Each individual smoothed penetration resistance curve $\chi_{s_i}(t)$ for $i = 1, \dots, 32$, was temporarily removed, and further predicted from the remaining ones by means of FKTM. Figure 6 shows the cross-validation predictions

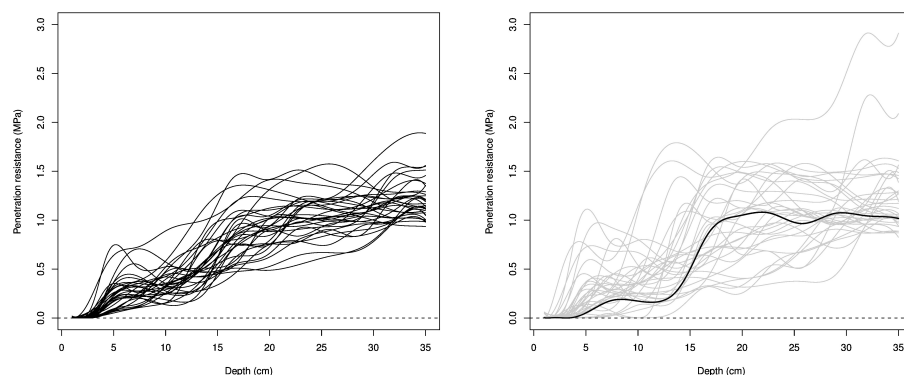


Figure 6: *Left panel: Cross-validation predictions of penetration resistance curves. Right panel: Predicted penetration resistance at a unvisited site obtained by FKTM (dark line) and smoothed penetration resistance curves (gray color). Source: own elaboration.*

(left panel), and the smoothed curves, and the prediction at an unvisited site (right panel). A graphical comparison between smoothed and predicted curves shows the good performance of the predictions (Figure 6). A comparison with OKFD and CTKFD in terms of the sum of squared errors (Table 3) shows small differences among the methods. A Friedman test (based on the SSE_F values) indicates that there is no significant difference (Friedman chi-squared = 2.6875, $df = 2$, p -value = 0.2609) among the three predictors. We can observe in Table 3 that with FKTM we obtain lower values of maximum and standard deviation of the sum of squared errors, that is, FKTM allows to obtain better cross-validation results in several sites. A detailed study of the sum of squared error site by site shows that FKTM is better than other in 14 sites (44 % of them).

6 Discussion and further research

We have shown how two predictors used in multivariable geostatistics can be extended naturally to the functional context. Our first predictor (cokriging based on curves) can be used with the same goal that classical multivariable cokriging (univariate prediction), but now considering curves instead of vectors as auxiliary information. This methodology could be an alternative in space-time modeling. The second proposed predictor (functional kriging: total model) allows predicting a whole curve at an unvisited site. The estimation for both proposed predictors is based on the use of basis function expansion. We have shown that when using this approach our two proposals coincide. Our second approach is also considered by Nerini et al. (2010). They propose the use of an orthonormal basis because in that case the predictor is simpler. The use of an orthonormal basis is computationally

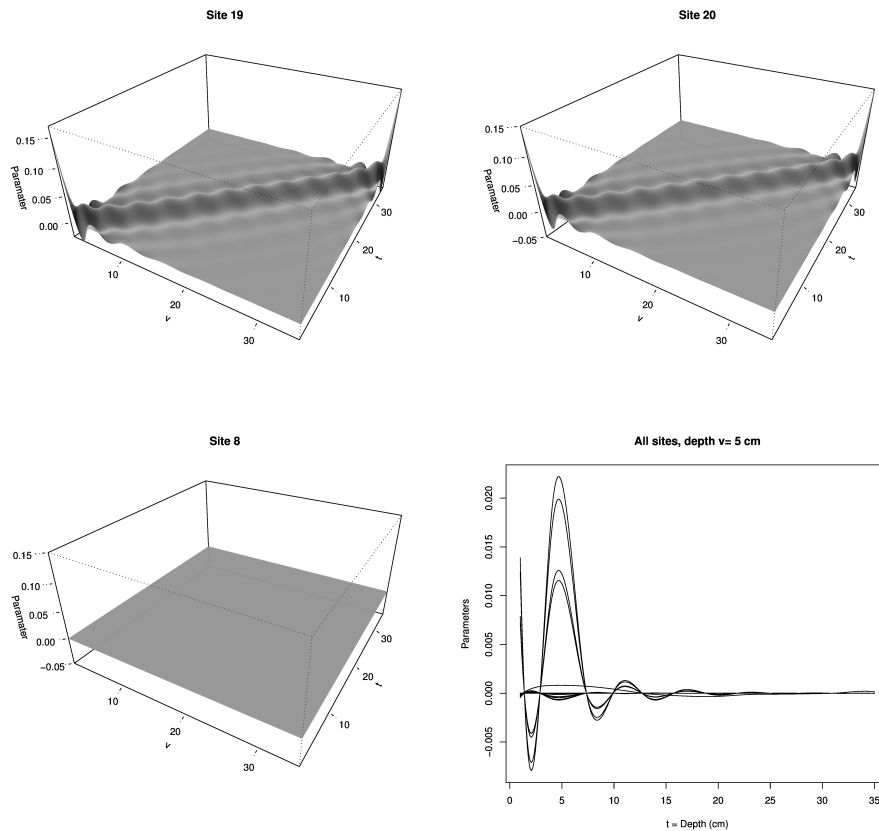


Figure 7: *Top: Estimated functional parameters for sites 19 and 20. Bottom: Estimated functional parameter for site 8 and functional parameters for all sites when we predict at an unvisited site on depth 5 cm. Source: own elaboration.*

Table 3: *Summary statistics of the sum of the squared errors of functional cross-validation with the penetration resistance data set. Results obtained when using B-Splines basis to smooth each curve. OKFD: Ordinary kriging for function-valued spatial data; CTKFD: Continuous time-varying kriging for functional data; FKTM: Functional kriging (total model), that equals CBFD (Cokriging based on functional data). Source: own elaboration.*

Statistic	OKFD	CTKFD	FKTM (or CBFD)
Minimum	4.3	8.2	4.8
Median	24.8	22.4	26.9
Mean	34.9	33.3	34.6
Maximum	175.8	202.2	172.5
Standard deviation	34.3	35.4	33.8
Sum	1118	1066	1108

advantageous but this is not a necessary condition for applying our proposals. Thus, other basis useful in FDA, such as B-splines, can be used.

The low temporal effect detected by the estimated parameters in the data analyzed suggests to consider only a domain $v - \delta \leq t \leq v + \delta$, where δ is a time lag over which we use information for prediction. The examples used suggest that the three predictors considered are equally useful. However, we think that further research is needed for establishing the performance of the proposed predictors under different levels of temporal cross-correlation.

Our proposals are based on both a common number of basis functions K and the same class of basis functions for both the functional variables and the parameters. We assume this for ease of mathematical developments. However an open question is if this could be generalized to using different K 's and types of basis functions. Another important point is that of relaxing the assumption of stationarity and isotropy.

A generalization to the case where the mean function change inside the study area is required. Models for carrying out spatial prediction based on information of several functional variables, that is, two or more functional variables observed at each sampling location, could also be considered.

Recibido: 16 de Junio de 2017

Aceptado: 28 de Septiembre de 2017

References

- Bogaert, P. (1996), 'Comparison of kriging techniques in a space-time context', *Mathematical Geology* **28**(1), 73–86.
- Borgault, G. & Marcotte, D. (1991), 'The multivariable variogram and its application to the linear coregionalization model', *Mathematical Geology* **23**(10), 899–928.
- Cardot, H., Crambes, C., Kneip, A. & Sarda, P. (2007), 'Smoothing spline estimators in functional linear regression with errors in variables', *Computational Statistics & Data Analysis* **51**(10), 4832–4848.
- Chan, K., Oates, A., Swan, A., Hayes, R., Dear, B. & Peoples, M. (2006), 'Agro-nomic consequences of tractor wheel compaction on a clay soil', *Soil & Tillage Research* **89**, 13–21.
- Cressie, N. (1993), *Statistic for spatial data*, Jhon Wiley & Sons, New York.
- Cuevas, A., Febrero, M. & Fraiman, R. (2004), 'An anova test for functional data.', *Computational Statistics & Data Analysis* **47**, 111–122.
- Ferraty, F. & Vieu, P. (2003), 'Curves discrimination. A non parametric functional approaches', *Computational Statistics & Data Analysis* **44**, 161–173.

- Ferraty, F. & Vieu, P. (2006), *Non parametric functional data analysis. Theory and practice*, Springer, New York.
- Freddi, O., Carvalho, M., Versonesi, M., Guilherme, J. & Carvalho, J. (2006), 'Relationship between maize yield and soil mechanical resistance to penetration under conventional tillage', *Engenharia Agrícola* **26**(1), 113–121.
- Galindo, J. (2004), Desarrollo de un modelo explicativo del crecimiento y producción de la arveja en función de variables edáfico-ambientales., PhD thesis, National University of Colombia.
- Giraldo, R. (2009), Geostatistical analysis of functional data, PhD thesis, Universitat Politècnica de Catalunya.
- Giraldo, R., Delicado, P. & Mateu, J. (2010), 'Continuous time-varying kriging for spatial prediction of functional data: An environmental application', *Journal of Agricultural, Biological, and Environmental Statistics* **15**(1), 66–82.
- Giraldo, R., Delicado, P. & Mateu, J. (2011), 'Ordinary kriging for function-valued spatial data', *Environmental and Ecological Statistics* **18**, 411–426.
- González-Manteiga, W. & Vieu, P. (2007), 'Statistics for functional data', *Computational Statistics and Data Analysis* **51**, 4788–4792.
- Goulard, M. & Voltz, M. (1993), Geostatistical interpolation of curves: A case study in soil science, in A. Soares, ed., 'Geostatistics Tróia '92', Vol. 2, Kluwer Academic Press, pp. 805–816.
- Hall, P. & Hosseini-Nassab, M. (2006), 'On properties of functional principal components analysis', *Journal of the Royal Statistical Society, Series B* **68**, 109–126.
- Hollander, T. & Wolfe, D. (1999), *Nonparametric Statistical Methods*, John Wiley & Sons, New York.
- Leiva, F. (2003), 'Manejo sostenible del suelo bajo la concepción de agricultura de precisión'. Research project partially supported by COLCIENCIAS, Colombia.
- Malfait, N. & Ramsay, J. (2003), 'The historical functional linear model', *The Canadian Journal of Statistics* **31**(2), 115–128.
- Myers, D. (1982), 'Matrix formulation of Co-kriging', *Mathematical Geology* **14**(3), 249–257.
- Nerini, D., Monestiez, P. & Manté, C. (2010), 'Cokriging for spatial functional data', *Journal of Multivariate Analysis* **101**, 409–418.
- Pebesma, E. (2004), 'Multivariable geostatistics in S. The gstat package', *Computers & Geosciences* **30**, 683–691.

- Ramsay, J. & Silverman, B. (2005), *Functional data analysis. Second edition*, Springer, New York.
- Stanley, D. (2002), *Canada's Maritime provinces. First edition*, Lonely Planet Publications, Marybirnong.
- Ver Hoef, J. & Barry, R. (1998), 'Constructing and fitting models for cokriging and multivariable spatial prediction', *Journal of Statistical Planning and Inference* **69**, 275–294.
- Ver Hoef, J. & Cressie, N. (1993), 'Multivariable spatial prediction', *Mathematical Geology* **25**(2), 219–240.
- Wackernagel, H. (1995), *Multivariable geostatistics: An introduction with applications*, Springer-Verlag, Berlin.
- Yamanishi, Y. & Tanaka, Y. (2003), 'Geographically weighted functional multiple regression analysis: A numerical investigation.', *Journal of Japanese Society of Computational Statistics* **15**, 307–317.

Appendix: proof of proposition 1

Let us define the following sets of functions: $\mathcal{F} = \{\lambda : T \times T \rightarrow \mathbb{R}\}$,

$$\mathcal{L}_C = \left\{ \lambda \in \mathcal{F}^n : \lambda_i(t, v) = \mathbf{b}_i^T(v) \mathbf{B}(t), \mathbf{b}_i(v) \in \mathbb{R}^K, v \in T, t \in T, i \in \{1 \dots n\} \right\},$$

$$\mathcal{L}_F = \left\{ \lambda \in \mathcal{F}^n : \lambda_i(t, v) = \mathbf{B}^T(v) \mathbf{C}_i \mathbf{B}(t), \mathbf{C}_i \in \mathbb{R}^{K \times K}, v \in T, t \in T, i \in \{1 \dots n\} \right\}.$$

Observe that $\mathcal{L}_F \subsetneq \mathcal{L}_C$.

Let \mathcal{F}_1 be the set of functions from T to \mathbb{R} and let g be a generic function defined from $\mathcal{F}_1^n \times T$ to \mathbb{R} . Let \mathcal{R} be a generic subset of \mathcal{F}^n such that $\mathcal{L}_F \cap \mathcal{R} \neq \emptyset$. For $\mathcal{L} \in \{\mathcal{L}_C, \mathcal{L}_F\}$ we define the functionals

$$\Psi_{\mathcal{L}}^{(1)}(g) = \int_T \min_{\lambda \in \mathcal{L} \cap \mathcal{R}} g(\lambda_1(\cdot, v), \dots, \lambda_n(\cdot, v), v) dv,$$

$$\Psi_{\mathcal{L}}^{(2)}(g) = \min_{\lambda \in \mathcal{L} \cap \mathcal{R}} \int_T g(\lambda_1(\cdot, v), \dots, \lambda_n(\cdot, v), v) dv.$$

Observe that for all function g and for all $\lambda^F \in \mathcal{L}_F \cap \mathcal{R}$,

$$\Psi_{\mathcal{L}_C}^{(1)}(g) \leq \Psi_{\mathcal{L}_F}^{(1)}(g) \leq \Psi_{\mathcal{L}_F}^{(2)}(g) \leq \int_T g(\lambda_1^F(\cdot, v), \dots, \lambda_n^F(\cdot, v), v) dv,$$

and

$$\Psi_{\mathcal{L}_C}^{(1)}(g) \leq \Psi_{\mathcal{L}_C}^{(2)}(g) \leq \Psi_{\mathcal{L}_F}^{(2)}(g) \leq \int_T g(\lambda_1^F(\cdot, v), \dots, \lambda_n^F(\cdot, v), v) dv.$$

Let λ^{C*} be the optimal element in $\mathcal{L}_C \cap \mathcal{R}$ in the sense that

$$\min_{\lambda \in \mathcal{L}_C \cap \mathcal{R}} g(\lambda_1(\cdot, v), \dots, \lambda_n(\cdot, v), v) = g(\lambda_1^{C*}(\cdot, v), \dots, \lambda_n^{C*}(\cdot, v), v),$$

or equivalently

$$\int_T g(\lambda_1^{C*}(\cdot, v), \dots, \lambda_n^{C*}(\cdot, v), v) dv = \Psi_{\mathcal{L}_C}^{(1)}(g).$$

If it would be possible to prove that in fact λ^{C*} is an element of $\mathcal{L}_F \cap \mathcal{R}$, then it would follow that

$$\int_T g(\lambda_1^{C*}(\cdot, v), \dots, \lambda_n^{C*}(\cdot, v), v) dv = \Psi_{\mathcal{L}_C}^{(1)}(g) \leq \Psi_{\mathcal{L}_F}^{(2)}(g) \leq \int_T g(\lambda_1^{C*}(\cdot, v), \dots, \lambda_n^{C*}(\cdot, v), v) dv$$

and then the inequalities would be in fact equalities. Therefore we would conclude that the minimums defining $\Psi_{\mathcal{L}_C}^{(1)}(g)$ and $\Psi_{\mathcal{L}_F}^{(2)}(g)$ are achieved at the same array of functions λ^{C*} .

We apply this result to the function g defined as

$$g(\lambda_1(\cdot, v), \dots, \lambda_n(\cdot, v), v) = \mathbb{V} \left(\sum_{i=1}^n \int_T \lambda_i(t, v) \chi_{s_i}(t) dt - \chi_{s_0}(v) \right)$$

and the subset \mathcal{R} of unbiased predictors,

$$\mathcal{R} = \left\{ \lambda \in \mathcal{F}^n : \mathbb{E} \left(\sum_{i=1}^n \int_T \lambda_i(t, v) \chi_{s_i}(t) dt \right) = \mathbb{E}(\chi_{s_0}(v)), \text{ for all } v \in T \right\}.$$

Then we only need to prove that the optimal coefficients $\lambda_i(v, t) = \lambda_i^v(t)$ defined in equation (8) as $\mathbf{b}_{iv}^T \mathbf{B}(t)$, are in fact of the form $\mathbf{B}^T(v) \mathbf{C}_i \mathbf{B}(t)$. So we need to prove that the optimal \mathbf{b}_{iv} are

$$\hat{\mathbf{b}}_{iv} = \hat{\mathbf{C}}_i \mathbf{B}(v)$$

for some matrix $\hat{\mathbf{C}}_i \in \mathbf{R}^{K \times K}$. But this is true from equations (18), (23) and (24) which establish that

$$\left(\mathbf{b}_{1v}^T, \dots, \mathbf{b}_{nv}^T, \mathbf{m}_v^T \right)^T = \mathbf{M}^{-1} \mathbf{N}(v) = \mathbf{M}^{-1} \begin{pmatrix} \mathbf{W} \mathbf{C}(\mathbf{a}_0, \mathbf{a}_1) \\ \vdots \\ \mathbf{W} \mathbf{C}(\mathbf{a}_0, \mathbf{a}_n) \\ \mathbf{W}^{-1} \end{pmatrix} \mathbf{B}(v),$$

and the proof concludes.